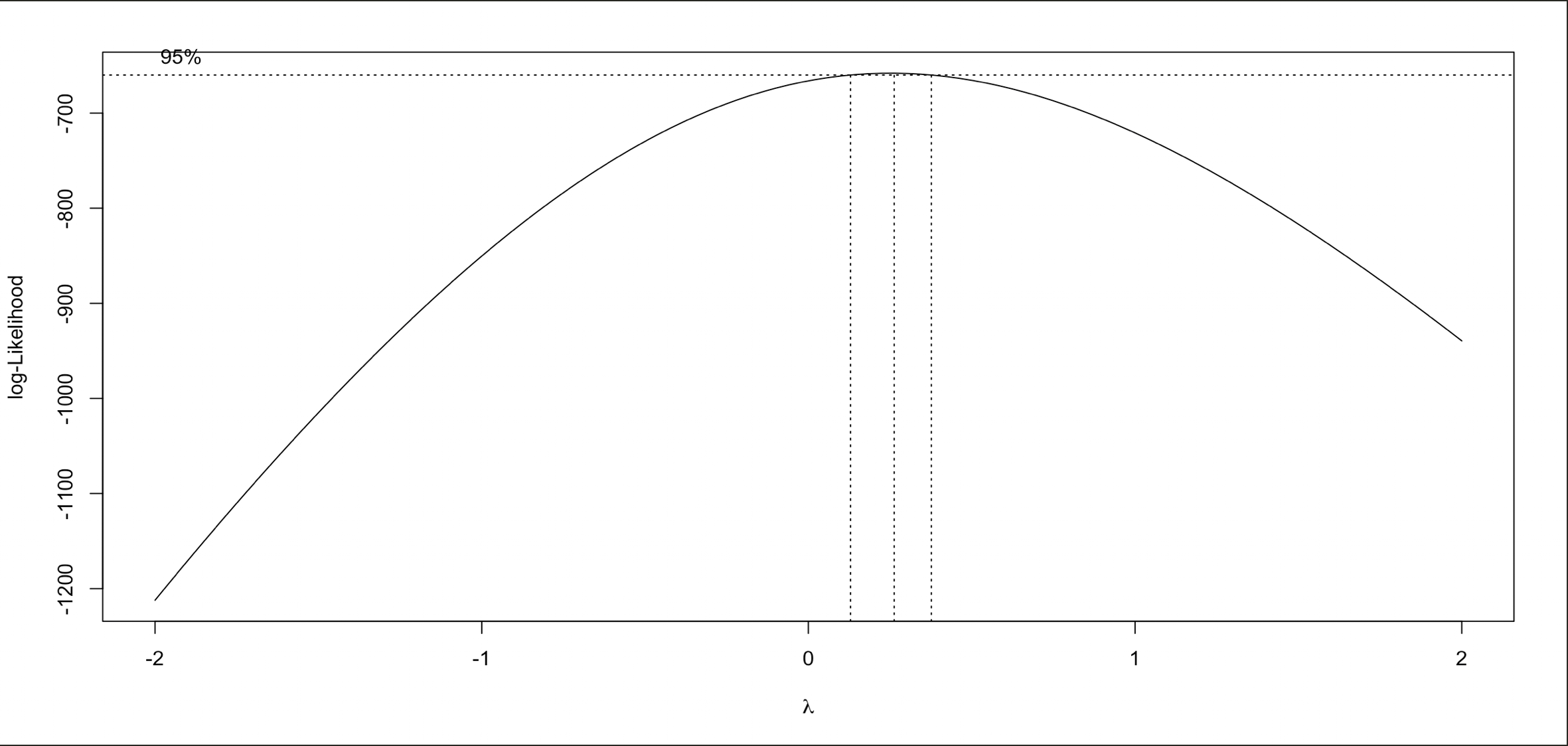
1. List all the points (row numbers) you removed (indexed on the original dataset) as outlier points.

* The points I removed are: 369, 372, 373, 370, 366, 365, 381, 419, 406, 411

1. Box-Cox Transformation - Plot the Box-Cox transformation curve (Log-likelihood vs Parameter value). What is the best value of the parameter you got?

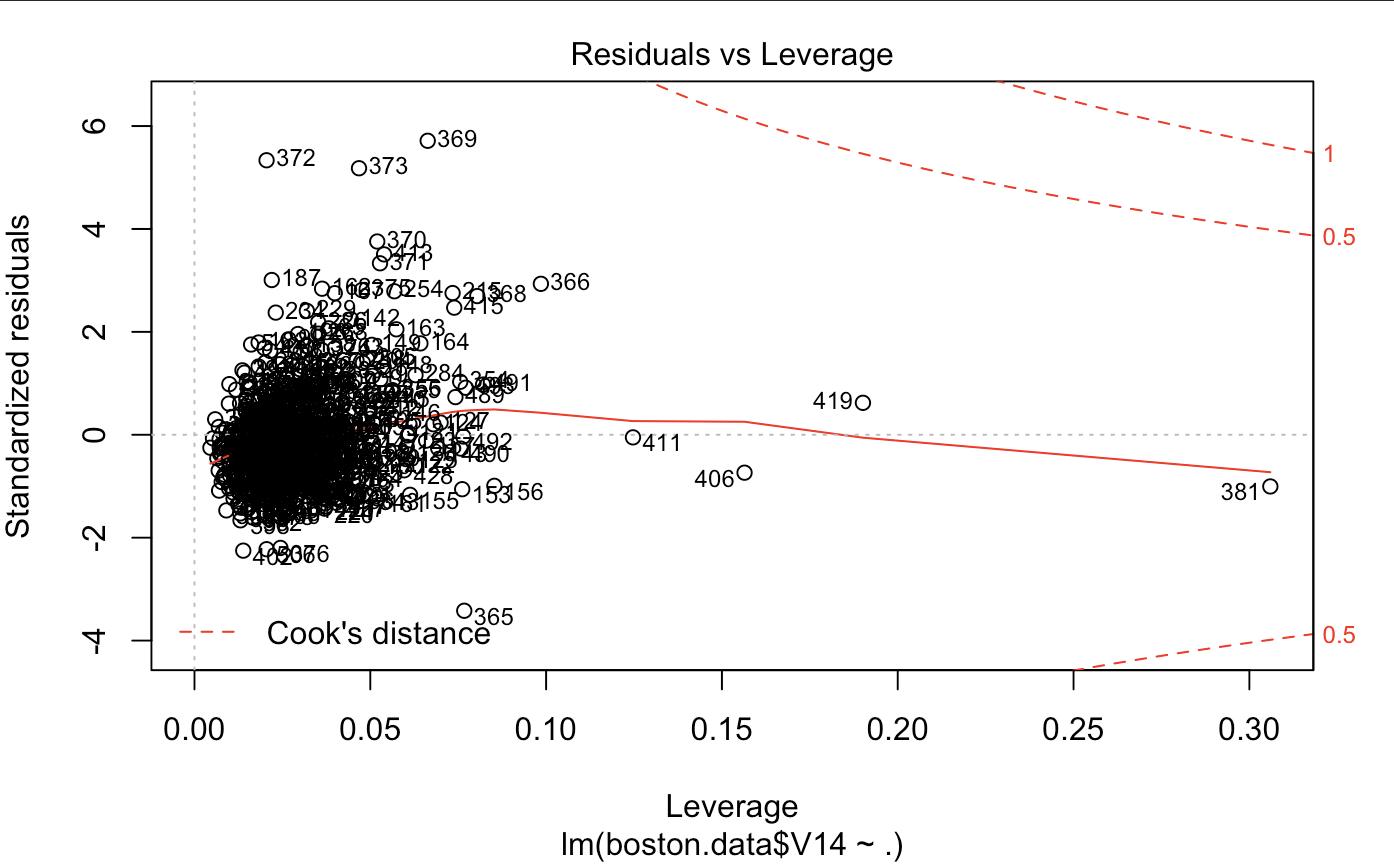
* The best value of parameter I got is 0.2626263
* Box-Cox transformation curve:



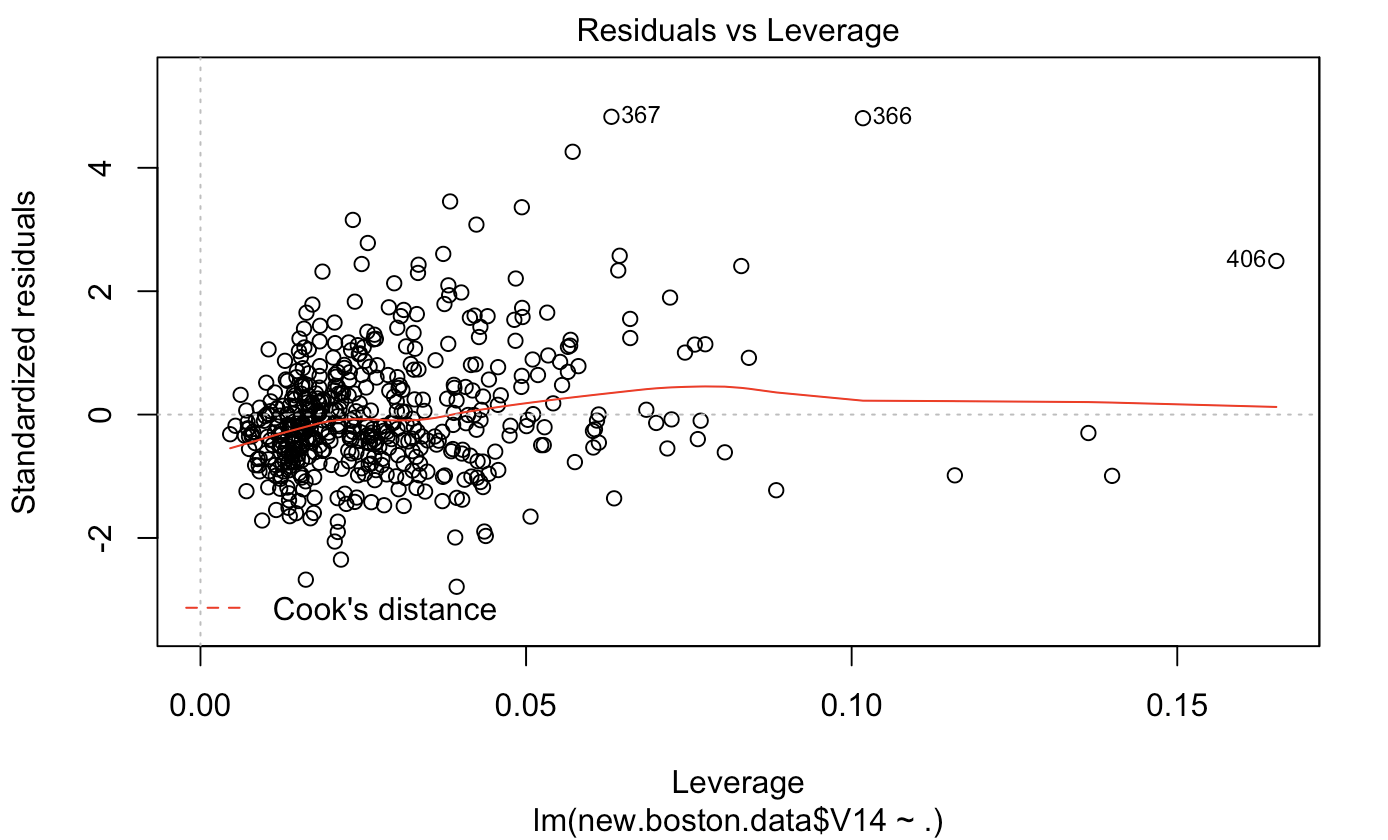
1. Diagnostic plots used for identification of outliers. Please only include the Standard residuals vs Leverage vs Cook’s distance plots (do not put other 3 plots you obtain for R). The final diagnostic plot obtained after removing all outliers should also be included.

* Standard residuals vs Leverage vs Cook’s distance plots without removing outliers

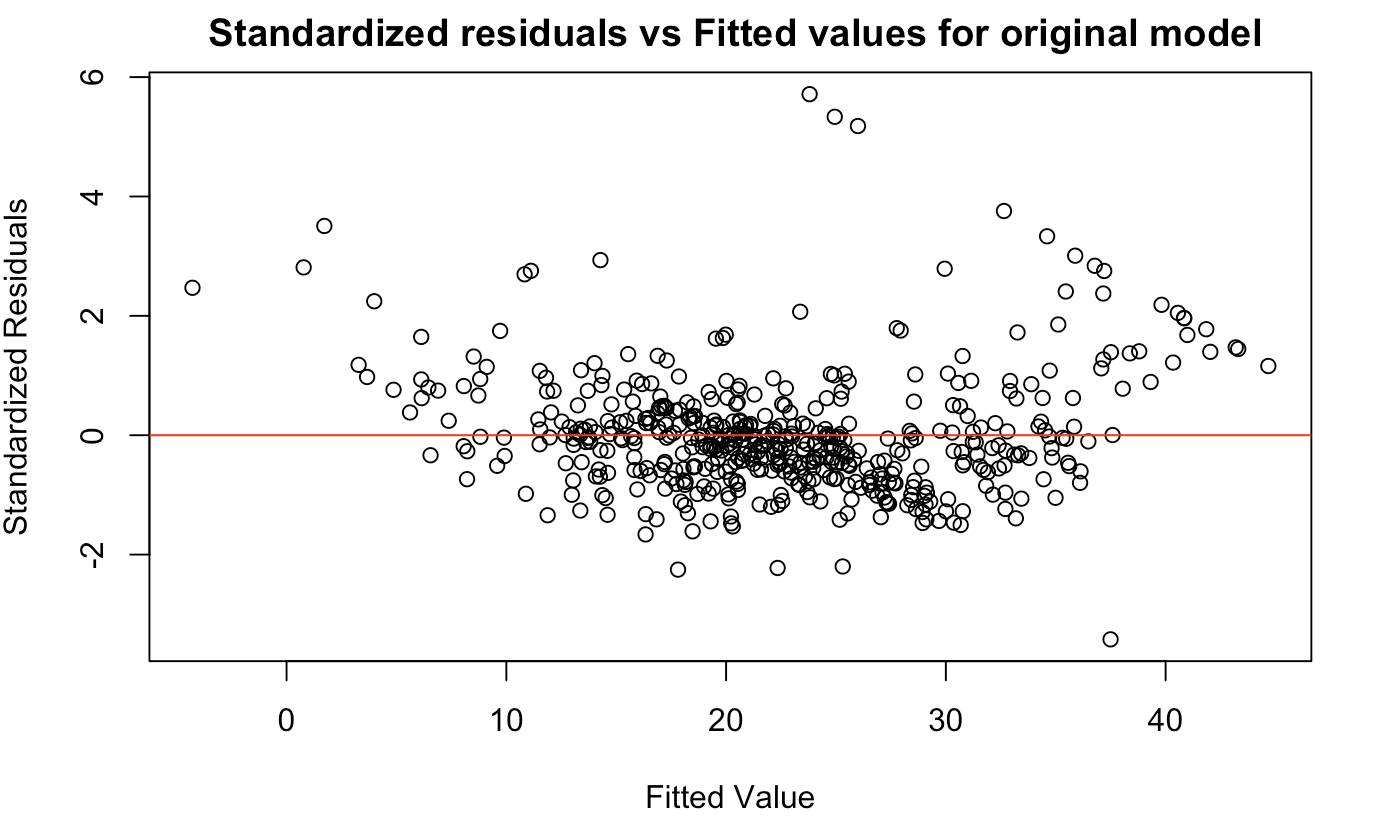
(I use “plot(model, which = 5, id.n=num.points)” to specify all the index of the points)



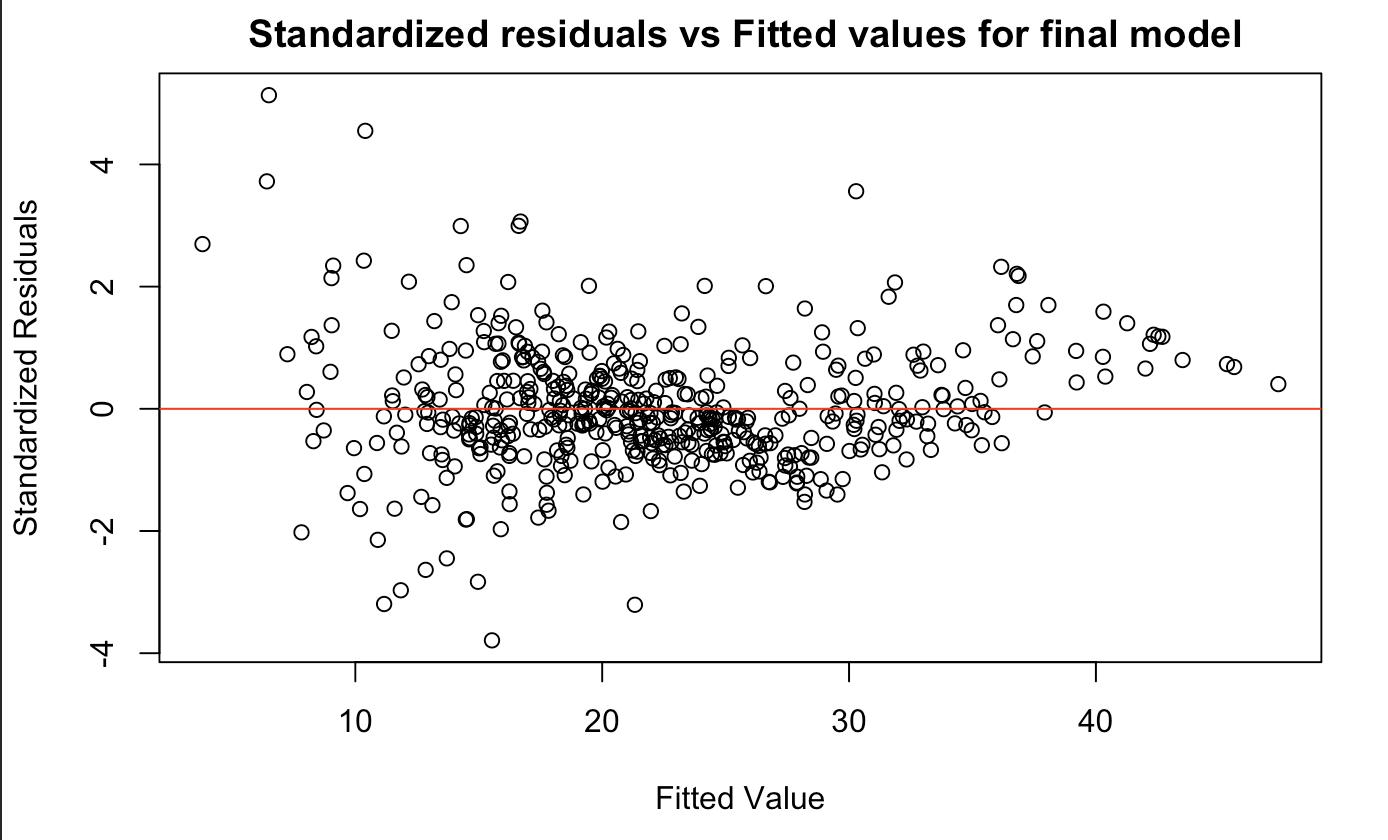
* Standard residuals vs Leverage vs Cook’s distance plots after removing outliers



1. Plot of Standardized residuals vs Fitted values for the linear regression model obtained without any transforms (like removing outliers or transforming dependent variables.



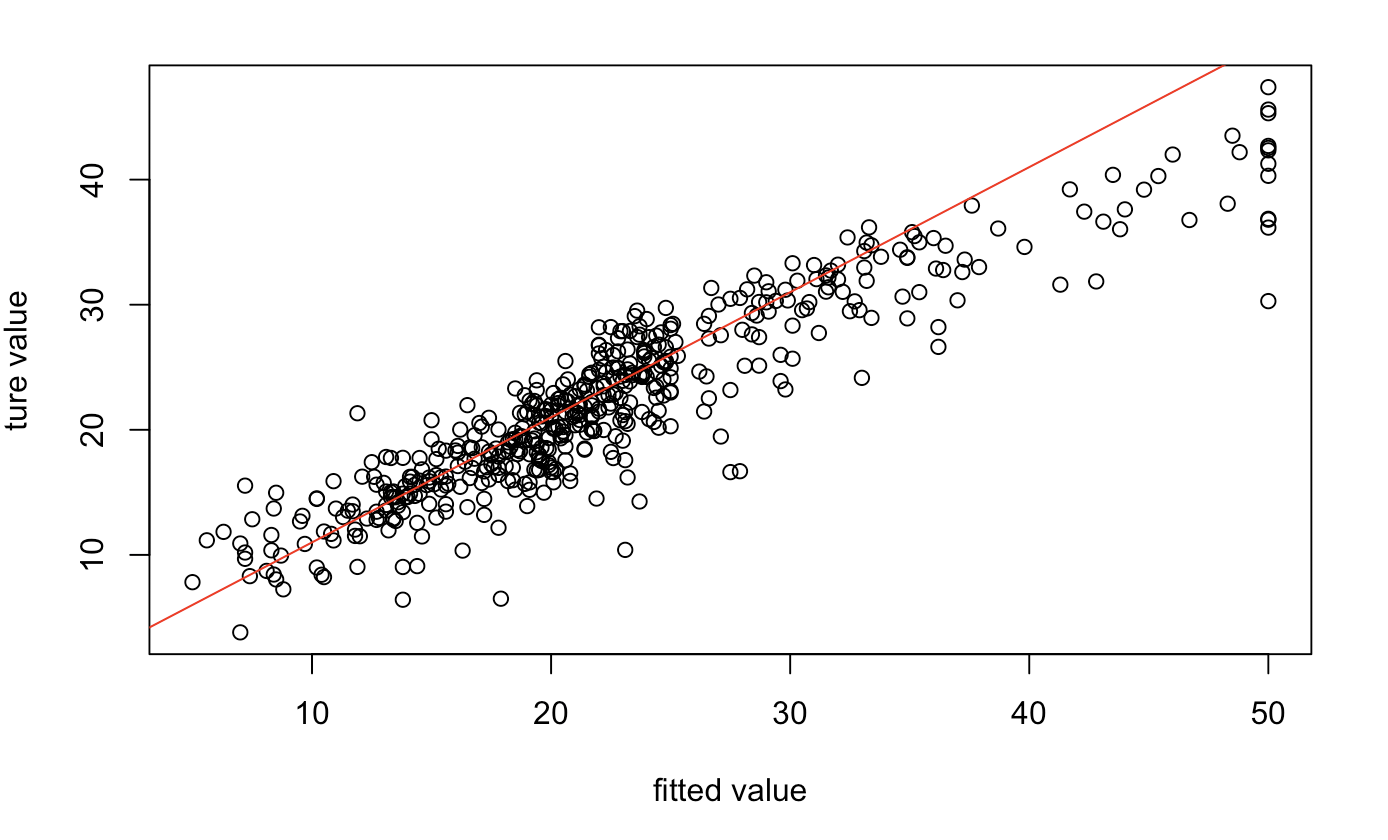
1. Plot of Standardized residuals vs Fitted values for the final linear regression model obtained after removing all outliers and transforming the dependent variable



1. Compare the two plots. What do you observe?

We observed that with the transformation and removing outliers, we get a better plot in the sense that the standardized residuals are more aligned to 0 and the residual seems to be uncorrelated to the predicted (which is a good sign) compared to the plot for the original model which has some non-linear relationship between the fitted values and the standardized residuals which suggests non-linear transformation will be helpful.

1. Final plot of Fitted house price vs True house price. What do you observe?



The relationship between the true house price and the predicted house price is mostly linear and follows the line y = x well although there exist some inaccuracies.

1. 1 page Code screenshot. It should include code for Linear regression, Box-Cox transformation and how you used the parameter value to transform the dependent variable.

